

## QUESTIONS FOR ADVISORY PANEL DISCUSSION

### **INTRODUCTION**

The PMA device is a first of a kind in the United States (i.e., total hip system with a resurfacing femoral component and metal-on-metal articulating surfaces). The PMA is supported by clinical data essentially from one source: the surgical experience of Dr. Derek J.W. McMinn, FRCS, who performed his surgeries at the Birmingham Nuffield Hospital, Edgbaston, Birmingham, United Kingdom. The PMA includes safety and effectiveness data from an uncontrolled, consecutive case series of all 2,385 procedures implanted with the Birmingham Hip Resurfacing (BHR) device by Dr. McMinn from July 1997 through May 2004. FDA requests expert clinical opinion regarding the safety and effectiveness data collection methods, the applicability of the foreign data from a single investigator and United Kingdom practice of medicine to the target United States population and practice of medicine, and the study results with respect to the device's safety and effectiveness.

### **STUDY OBJECTIVES AND ASSESSMENTS**

The objective of this PMA is to demonstrate the safety and effectiveness of the Birmingham Hip Resurfacing (BHR) System. The safety assessments included data on revisions, adverse events, deaths and a metal ion literature review. The primary effectiveness assessments included survivorship and radiographic data. The secondary effectiveness assessments included pain and function data as evaluated by the Oswestry-modified Harris Hip (OSHIP) Score, and patient satisfaction data.

### **DESCRIPTION OF COHORTS AND DATA COLLECTED**

The 2,385 procedures implanted with the Birmingham Hip Resurfacing (BHR) device by Dr. McMinn from July 1997 through May 2004 were divided into the following three main cohorts:

- **X-ray cohort:** First 124 BHR cases performed by McMinn from July 1997 through December 1997.
- **Oswestry cohort:** Next 1502 BHR cases performed by McMinn from January 1998 through March 2002.
- **McMinn cohort:** Next 759 BHR cases performed by McMinn from April 2002 through May 2004.

The following table outlines the dates of implantation, number of procedures, and types of safety and effectiveness data collected for these 3 cohorts:

			Types of Safety and Effectiveness Data Collected						
			Safety Data Collected			Effectiveness Data Collected			
Cohort	Dates of Implantation	Number of Procedures	Adverse Events	Revisions	Deaths	Survivorship	Radiographic	Pain and Function (OSHIP)	Patient Satisfaction
X-ray	7/97-12/97	124	X	X	X	X	<b>X</b>	X**	X
Oswestry	1/98-3/02	1502	X	X	X	X		X**	X
McMinn	4/02-5/04*	759*	X	X	X	X		***	

Note: An X in the table indicates that this data was collected for the respective cohort

\* The sponsor stated that there were 5 cases in the McMinn cohort whose implantations were performed prior to 4/02. These cases should have been part of the Oswestry cohort, but for unknown reasons were not. Therefore, unlike the majority of the McMinn cohort, these 5 cases have longer term follow-up.

\*\* See note in Table of combined cohorts below regarding the number of procedures contributing to the pain and function (OSHIP) effectiveness data.

\*\*\* The pain and function data for the procedures in the McMinn cohort were collected using the Oxford Hip Score evaluation method (and not the OSHIP Score). The sponsor explained that because the 759 procedures in the McMinn Cohort were not tracked by the Oswestry Outcome Center but by the National Health Services (NHS) Center, the sponsor did not have access to the Oxford hip score data.

As noted in the Table above (with the large bolded “X”), only 124 procedures in the X-ray cohort contributed to the assessment of radiographic effectiveness in the PMA. Radiographic evaluations were not provided for the 1502 procedures in the Oswestry cohort or the 759 procedures in the McMinn cohort.

Where there were common data elements collected in the 3 cohorts outlined above, the sponsor pooled this information into the following two combined cohorts:

- **X-ray/Oswestry/McMinn combined cohort or Overall McMinn cohort:** Note that for the rest of this Executive Clinical Summary, this cohort will be referred to as the **Overall McMinn cohort**.
- **X-ray/Oswestry combined cohort**

The following table outlines the dates of implantation, number of procedures, and types of safety and effectiveness data collected for these 2 combined cohorts:

			Types of Safety and Effectiveness Data Collected						
			Safety Data Collected			Effectiveness Data Collected			
Cohort	Dates of Implantation	Number of Procedures	Adverse Events	Revisions	Deaths	Survivorship	Radiographic	Pain and Function (OSHIP)	Patient Satisfaction
Overall McMinn Cohort	7/97-5/04	2,385	<b>X</b>	<b>X</b>	<b>X</b>	X			
X-ray/Oswestry Combined	7/97-3/02	1,626	X	X	X	<b>X</b>		<b>X**</b>	<b>X</b>

Note: An X in the table indicates that this data was collected for the respective cohort

\*\* Only 1,111 unilateral procedures in the X-ray/Oswestry combined cohort contributed to the assessment of pain and function effectiveness data, as evaluated by the Oswestry-modified Harris Hip (OSHIP) Score.

As noted in the Table above (with large bolded “X”s), the 2,385 procedures in the Overall McMinn cohort contributed to the assessment of safety including adverse events, revisions, and deaths. The 1,626 procedures in the X-ray/Oswestry combined cohort contributed to the assessment of survivorship. Also, as noted in the Table above, only 1,111 unilateral procedures in the X-ray/Oswestry combined cohort contributed to the assessment of pain and function effectiveness data, as evaluated by the Oswestry-modified Harris Hip (OSHIP) Score. See discussion below regarding unilateral and bilateral involvement when evaluating pain and function. Finally, 1,626 procedures in the X-ray/Oswestry Combined cohort contributed to the patient satisfaction effectiveness in the PMA.

#### **Additional Data Sources:**

The main data sources were presented above but the sponsor also included additional, less complete data on 3,374 BHR cases performed by 140 surgeons worldwide (other than Dr. McMinn). The follow-up for these cases was also contracted to the Oswestry Outcomes Centre and includes primarily the same parameters as the follow up for the X-ray/Oswestry combined cohort (adverse events, revisions, deaths, pain and function (OSHIP) scores, and patient satisfaction. The Oswestry Outcomes Centre, therefore, collected data on a total of 5000 BHR cases. These 5000 cases are referred to as the **Oswestry Worldwide Cohort**. The Oswestry Worldwide Cohort consists of 1) the 1626 McMinn cases of the X-ray/Oswestry cohort, and 2) an additional 3,374 non-McMinn (“all other”) cases. The Oswestry Outcomes Centre has provided Smith & Nephew access to all available data for the BHR cases from its database. Although the sponsor considers the data from the 3,374 “all other” cohort to be of some value, Smith & Nephew has no ability to independently verify any of the data provided to the Oswestry Outcomes Centre by sites other than the McMinn Center, and has no ability to request additional follow-up or clarifications of any kind from non-McMinn patients or physicians. For these reasons, the analysis on the Oswestry Outcomes Centre worldwide database has some limitations, and is not considered a primary data source for this PMA.

## **Outline of Items for discussion:**

---

- **Data Collection Methods**
    - **Safety Data Collection Methods (Question #1)**
    - **Effectiveness Data Collection Methods (Question #2)**
  - **Applicability of the Foreign Data from a Single Investigator and United Kingdom Practice of Medicine to the Target United States Population and Practice of Medicine (Question #3)**
  - **Device Safety (Question #4)**
  - **Device Effectiveness (Question #5)**
  - **Labeling (Question #6)**
  - **Post-Approval Study (Question #7)**
- 

## **DATA COLLECTION METHODS**

### **Safety Data Collection Methods**

The safety data including adverse events, revisions, and deaths were collected by:

- The Oswestry Outcomes Center using an annual, patient-completed, mail-in questionnaire (deaths were identified while attempting to perform scheduled follow-up);
- The McMinn Center by recording the findings of post-operative patient visits to the McMinn Center in patient records; and
- Recording information provided to Dr. McMinn by primary care physicians.

Dr. McMinn's follow-up was described as follows:

- Dr. McMinn performed regular evaluations (history, physical examination, radiographs to assess implant status, and any necessary laboratory work) in the preoperative and postoperative time periods according to standard practice, although the timepoints and evaluations were not according to a standard protocol.
- All revision surgeries were performed by Dr. McMinn (except in one known case). Therefore, the revision status was directly known to Dr. McMinn.
- There were no pre-defined follow-up time windows, standardized clinical evaluations, adverse event report forms, or standardized radiographic evaluations.

In addition, the Oswestry Outcome Center (OOC) provided the following information regarding follow-up procedures:

- The Oswestry Outcomes Center collected data on revisions and adverse events using an annual, patient-completed, mail-in questionnaire.
- With the exception of 8 cases classified by OOC as “no consent” (subjects who withdrew or did not agree to participate in the study), all other cases are not considered lost-to-follow-up by OOC since they continue to make attempts to contact patients.
- Of the 180 cases missing their last theoretical expected mail-in questionnaire follow-up, 84 are missing at least 2 yearly evaluations, while 96 are only missing

their last evaluation. These cases represent only 11.1% (180/1626) of the cases in the Oswestry/X-Ray Combined Cohort

- The steps taken at the OOC to regain contact if a patient does not respond to a request for information, are as follows:
  - Send reminder letters;
  - Use e-mail with request for data;
  - Contact consultant surgeon by letter for patient's current contact details;
  - Use the National Strategic Tracing Service database to determine patient whereabouts;
  - Contact surgeon and/or patient by telephone;
  - Attempt trace using online (Internet) census information;
  - If none of these produce results, an annual request for an update on progress is sent to the patient's last known address by Royal Mail and e-mail until the tenth anniversary of the operation. Patients are not classified as lost-to-follow-up until all avenues have been exhausted.

The sponsor states that they performed a 100% audit of all 2,385 procedures in the Overall McMinn Cohort, and therefore believe that all reported adverse event information have been captured.

In addition to the safety data collection methods outlined above, the sponsor provided a metal ion literature analysis. Included in the sponsor's analysis was an unpublished report by Daniel J, Ziaee H, and McMinn D, entitled, "Metal ion studies in patients treated with the Birmingham Hip Resurfacing, a comparable FDA-approved device and historic metal-metal total hip replacements." The authors conducted 4 metal ion studies in patients who received BHR, Metasul metal-metal total hip replacements, and other (historic) metal-metal total hip replacements. The four studies included:

1. A short-term longitudinal study of urinary Co and Cr levels in patients with the BHR and the Metasul metal-metal total hip replacement.
2. A long-term cross-sectional study of urinary Co and Cr levels in patients with the BHR and the Metasul metal-metal total hip replacement.
3. A longitudinal study of whole blood Co and Cr levels in patients with the BHR.
4. A cross-sectional study of whole blood Co and Cr levels in patients with the BHR and metal-metal total hip replacements.

In addition, the sponsor provided a summary of 18 literature references pertaining to the medium and long-term safety of cobalt and chromium ion exposure in the subject device (BHR), metal-on-metal total hip replacements, and metal-on-polyethylene total hip replacements. These literature references are summarized in the Executive Clinical Data Summary – Summary of Safety Data.

**Question #1:**

**Please discuss the evaluation methods used to collect safety data (i.e., data on revisions, adverse events, deaths, metal ion literature analysis) and whether or not these methods are reliable to assess the safety of the device.**

## **Effectiveness Data Collection Methods**

### **Primary Effectiveness Data Collection Methods**

#### **Survivorship Data Collection Method:**

The primary effectiveness measurement was the X-Ray/Oswestry Combined Cohort survivorship study that included 1626 procedures performed from 7/97 through 3/02 at the Birmingham Nuffield Hospital. These procedures were a minimum of 2 years post-op. Of the 1626 procedures, data are available for 546 of the 601 BHR procedures eligible for 5 year follow up (90.8%). The data for the survivorship study was collected using the same methods presented above for the safety data collection methods.

#### **Radiographic Data Collection Method:**

The PMA contained the results of an independent radiographic review of the X-Ray Cohort, the first 124 procedures performed in the series from 7/97 through 12/97. Radiographic evaluations were not provided for the 1502 procedures in the Oswestry Cohort or the 759 procedures in the McMinn Cohort.

Radiographs were taken on 108 of the 118 procedures expected at 5 years postoperatively (91.5%). Six (6) procedures were not expected at 5 years postoperatively because one patient with bilateral hip implants died from a motor neuron disease unrelated to the BHR procedure; and 4 of the 124 BHR procedures (3.2%) have undergone revision: 3 cases were revised for infection, and 1 case required revision because of a femoral neck fracture. Therefore, 118 procedures (124 hips - 2 hips due to death - 4 revisions = 118 procedures) were eligible for 5 year radiographic evaluation of the BHR. Ten other cases were missing due to lost to follow-up or incomplete film records. Therefore, one hundred and eight (108) of the 118 hips surviving to 5 years had 5 year radiographs available for independent review (91.5%). (Note: The sponsor reported that an additional bilateral patient died 7 years post-op due to stroke but had 5 year x-rays taken).

There were immediate post-operative films on 89 of the 108 procedures with 5-year radiographs but the sponsor stated that these films were low quality portable films and unusable for the purposes of precise postoperative measurement comparisons. Therefore, baseline films for the purposes of comparisons were made in each of the 108 cases in the postoperative time period (usually within 3 months, but 8 of the 108 procedures had baseline evaluations performed at time points ranging from 110-860 days).

The radiographs were interpreted by Dr. Nick Evans (Royal Orthopaedic Hospital, Birmingham, UK). The sponsor states that a prospective protocol was used to assess the radiographs. The 5-year AP and lateral view radiographs were compared with the baseline radiographs for the following:

- Medial-lateral migration (reference point = Kohler's line)
- Acetabular orientation (tilt angle)

- Femoral and acetabular radiolucencies (femoral: Amstutz defined 1-3 zones; acetabular: DeLee and Charnley defined I-III zones), each graded on a 0-9 scale. Radiolucency is defined as a lucent area parallel to and in close proximity to the prosthesis/bone interface encompassing at least 50% of the zone and at least 1mm in width.
- Heterotopic ossification (HO) (Brooker classification, I-IV)
- Other radiographic findings, including bone resorption, acetabular protrusion, cysts, buttressing, and other abnormalities.

A radiographic success was defined as having all of the following:

- Absence of radiolucencies or a radiolucency in any one or two zones (a score of 0-6);
- Component migration  $\leq 2\text{mm}$ ; and
- Change in acetabular angle  $< 5^\circ$

A radiographic failure is defined as the following:

- Presence of incomplete or complete radiolucencies or a radiolucency in all zones (a score of 7 or 8);
- A migration of the component  $> 2\text{mm}$ ; or
- A change in acetabular orientation of  $\geq 5^\circ$

The individual success criterion was the absence of radiographic findings that suggest revision is necessary.

## **Secondary Effectiveness Data Collection Methods**

### Oswestry-Modified Harris Hip (OSHIP) Score Data Collection Method

The clinical data used to support this PMA were collected by the Oswestry Outcomes Center using an annual, patient-completed, mail-in questionnaire. The responses to the pain, function, and movement questions in the questionnaire were used to generate the Oswestry-modified Harris Hip (OSHIP) Score.

The OSHIP questionnaire allows patient assessments without direct physician or examiner evaluation. No other sources of pain and function information were used to support this PMA.

Data entry for the OSHIP outcomes data used to support this PMA was performed by the trained employees of the Oswestry Outcomes Center. The Center's standard operating procedures for data input and clarification for the patient-administered OSHIP questionnaires were as follows:

- Trained employees carefully reviewed all questionnaires and identified any unclear, incomplete, or ambiguous items by highlighting them. Available information was recorded in the database, and any missing/unclear/conflicting information was recorded as "missing data" (indicated by an asterisk).
- The questionnaires with missing data were returned to the patients with specific instructions on how to complete or clarify any suspect data fields. The preferred

method of follow-up for missing data was by mail; however, secondary methods of follow-up included e-mail and telephone.

- Upon receipt of missing data, the trained employees verified that the highlighted data fields had been completed/clarified, and the data was entered into the database.
- All reasonable attempts to collect the data via mail, e-mail, or telephone were made. If all reasonable efforts at data collection failed, the score for any missing item was assumed to be the lowest possible (typically zero), unless otherwise indicated.

The sponsor stated that in an unpublished paper titled, “A Self-completed Tool for Evaluation of Hip Function: The Oswestry Hip Score,” D. Barnes and co-workers reported that the OSHIP was developed by Professor James Richardson FRCS (Orth), Professor of Orthopaedics at the Institute of Orthopaedics, Robert Jones and Agnes Hunt Orthopaedics and District Hospital—NHS Trust in Oswestry, Shropshire, England. According to Barnes’ paper, creation of the OSHIP began with the following premises:

- Long-term evaluation following hip replacement is essential and follow-up must be regular.
- Large-samples are necessary.
- Long-term and large-sample follow-up is difficult to obtain when using a score that requires surgeon- or radiologist-assessment.
- Physician-administered surveys are susceptible to bias (which may inflate the final scores) and may not truly represent the patients’ own feelings.
- Existing patient-completed scores lack accurate measurements of range of movement.
- Questionnaires need to be simple and relatively short to make long-term and large-scale collection of data more efficient.

Building on these premises, Professor Richardson developed the OSHIP by combining elements of both the Harris and Merle d’Aubigne scores. The OSHIP produces an overall index score similar to that of the Harris score between 0 (worst) and 100 (best). Both the OSHIP and Harris Hip Score (HHS) are made up of the three domains of pain, function, and hip movement, with function being further divided into gait (walking, limp, and distance), and activity (stairs, sitting and transport).

The main difference between the OSHIP questionnaire and the HHS is that the OSHIP allows patient assessments without direct physician or examiner evaluation. In addition, the OSHIP questionnaire does not include the three HHS questions regarding physician assessment of Range of Motion (5 pts.), Absence of Deformity (4 pts.), and the patient’s ability to put on socks/tie shoes (4 pts.) but substitutes a “movement” question (13 pts.) that is intended for the patient to estimate their ability to flex their hip.

The sponsor provided references to justify the use of the OSHIP in lieu of the HHS and the validity of self-administered questionnaires.



While Ragab<sup>1</sup> reported a lack of correlation between patient self-assessment of pain/function and physician assessment of pain/function ( $r=0.467$ ,  $p<0.01$ ), several others have reported the opposite—a very close correlation between patient self-assessment and physician assessment.

Research by Mahomed et al.<sup>2</sup> demonstrated that patients are able to accurately respond to HHS questions regarding pain and function with little difficulty, and that there is excellent correlation between the overall HHS pain/function scores reported by patients and the overall HHS pain/function scores reported by physicians (Pearson correlation coefficient=0.99,  $p<0.0001$ ). In this study, Mahomed also reported that the  $\kappa$  statistic, which is a measure of the reproducibility between repeated assessments of the same categorical variable, ranged between 0.79 and 1.00 ( $p<0.0001$ ) for each item of the HHS and, according to the paper, “indicated excellent reproducibility.” Note that both Ragab and Mahomed’s studies did not include patient or physician evaluations of range of motion or deformity, these questions were eliminated from both the patient and physician assessments. Furthermore, McGrory et al.<sup>3</sup> found that a brief follow-up phone call (similar to the OOC follow-up procedure discussed above) was effective in capturing missing data and clarifying multiple or contradictory responses from mailed patient self-assessment questionnaires.

Barnes et al. have evaluated the reliability and validity of the Oswestry Hip Score as documented in the research paper, “A Self-completed Tool for Evaluation of Hip Function: The Oswestry Hip Score.” In Mr. Barnes’ study, a group of 61 patients completed the Oswestry Hip Score (OSHIP). They were then sent a second copy of the OSHIP to be completed two weeks later and returned by mail. The results of these two sets of surveys were compared to look for reproducibility. When comparing the OSHIP responses from the first self-administration to the results of the second self-administration two weeks later, the total intra-class correlation coefficients was 0.93 with intra-class correlation coefficients for the individual items and domains ranging from 0.67 to 0.92.

Mr. Barnes’ study also included a separate group of 28 consecutive patients who were given both a patient-administered OSHIP and a physiotherapist-administered Harris Hip Score. The correlation between the patients’ overall self-administered OSHIP scores and physiotherapist-administered overall HHS scores was 0.91 ( $p<0.0001$ ). Correlation between the individual corresponding domains of the Oswestry Hip Score and Harris Hip Score ranged from 0.60 and 0.89. The strongest correlation was between the domains of ‘stairs’ and ‘walking/support’ (0.89) and the lowest for the domains of ‘limp’ (0.60). FDA requested additional correlations be provided that were not included

---

<sup>1</sup> Ragab A.A. *Validity of self-assessment outcome questionnaires: patient-physician discrepancy in outcome interpretation*, Biomed Sci Instrum (2003); 39: pp.579-84

<sup>2</sup> Mahomed NN, Arndt DC, McGrory BJ, Harris WH. *The Harris Hip Score: Comparison of patient self-report with surgeon assessment*. J Arthroplasty 16(5):575-80, 2001.

<sup>3</sup> McGrory BJ, Shinar AA, Freiberg AA, Harris WH. *Enhancement of the value of hip questionnaires by telephone follow-up evaluation*. J Arthroplasty 12(3), 1997.

in Mr. Barnes' study. Correlation between the OSHIP "movement" domain and the HHS "shoes & socks," "deformity," and "range of motion" domains were requested and performed by the sponsor. The correlation between OSHIP "movement" and HHS "shoes and socks," and HHS "range of motion," was 0.40 and 0.21, respectively. The sponsor stated that the correlation between OSHIP "movement" and HHS "deformity" was not included and not useful because all 28 subjects scored the maximum of 4 points on the HHS scale (score is either 0 or 4). FDA performed additional correlations between OSHIP "movement" domain and the sum of the scores for the HHS "range of motion," "shoes and socks," and "deformity." The correlation between these items was calculated because the OSHIP "movement" domain is the substitute for the HHS "range of motion," "shoes and socks," and "deformity" domains. The correlation was calculated to be 0.40. In addition, FDA performed a linear regression analysis to predict HHS total score from OSHIP total score for the 28 subjects. The linear regression analysis is summarized in the Executive Statistical Summary and the calculated  $R^2$  is approximately 0.83, which measures the proportion of total variation about the mean explained by the linear regression model (Fitted HHS =  $9.6 + (0.9276 \times \text{OSHIP})$ ). FDA believes that due to an unclear randomization scheme and questionable masking procedure used to select these 28 sample patients, it is not easy to generalize the above correlations to the general target patient population. Clinical judgement is needed.

Like the Barnes study, Ragab<sup>4</sup> also reported a relative lack of correlation between patient assessment of limp and physician assessment of limp which he believed was due to the physician's tendency not to report limps that occurred only after long walks or during weather change, while patients were likely to report such limps. However, unlike the Barnes study in which the OSHIP and HHS item regarding "pain" had a correlation of 0.83, Ragab found that when the patients reported significant pain (i.e., pain scores less than 30), they were often attributing the pain to their hips when the pain, in most cases, was not truly hip related. The author reported that the physician was better able to distinguish "true" hip pain from pain coming from other sources (e.g., secondary to trochanteric bursitis, lumbar spondylosis, arthrosis of the contralateral hip).

An additional finding by McGrory and co-workers<sup>5</sup> was that questions about whether patients could cut their toenails and put on socks/shoes correlated significantly with postoperative weighted HHS range of motion calculation ( $P = 0.0002$ ,  $r = 0.569$ ,  $R^2 = 0.323$  for cutting toe nails and  $P = 0.0006$ ,  $r = 0.529$ ,  $R^2 = 0.280$  for putting on socks and shoes). The authors concluded that responses to these two questions could therefore be used to estimate the weighted HHS range of motion. In addition, Johnston and Smidt<sup>6</sup> reported that there is a distinct relationship between hip flexion and shoe tying. In a motion study of 135 post-total-hip-replacement cases, they found that

---

<sup>4</sup> Ragab A.A. *Validity of self-assessment outcome questionnaires: patient-physician discrepancy in outcome interpretation*, Biomed Sci Instrum (2003); 39: pp.579-84.

<sup>5</sup> Johnson RC, Smidt GL. Hip motion measurements for selected activities of daily living. Clin Orthop 72:205-216, 1970.

<sup>6</sup> McGrory BJ, Freiberg A, Shinar AA, Harris WH. Correlation of measured range of motion following tota hip arthroplasty and responses to a questionnaire. J Arthroplasty 11(5):565-71, 1996.

patients with  $\geq 120^\circ$  of flexion could tie their shoes. They found that the majority of patients with  $90\text{-}100^\circ$  of flexion could tie their shoes with difficulty. They found that the majority of patients with  $<90^\circ$  of flexion could not tie the shoes.

A review of the raw data from the Barnes' study, as described above, of 28 patients given both a patient-administered OSHIP and a physiotherapist-administered Harris Hip Score also revealed the following:

- The average OSHIP score for the 28 subjects was 62.25, while the average HHS score was 67.36.
- The OSHIP results indicated that 9 of 28 (32%) subjects achieved an overall score of 80 or better, while the HHS results indicated that 12 of 28 (43%) of subjects achieved an overall score of 80 or better.
- The OSHIP results indicated that 14 of 28 (50%) subjects scored  $<70$ , while the HHS results indicated that 12 of 28 (43%) subjects scored  $<70$ .
- There were 14 pairs of data where the OSHIP and HHS scores differed by more than 5 points. Of the 14, the HHS score was higher in 12 cases (85.7%) while the OSHIP was higher in only 2 cases (14.3%).

In the final comment from the study by D. Barnes and co-workers, the authors stated that the Oswestry Hip Score is not intended to replace clinical examinations at the critical phases following hip surgery (i.e., 1-year, 5-years, and 10-years). However, it can be a useful tool along with X-rays to replace unnecessary yearly follow-up following hip surgery.

The sponsor used the referenced studies by Mahomed, McGrory, and Barnes to justify the use of patient self-administered questionnaires to adequately report pain and function data. Furthermore, the sponsor asserted that the close correlation of the overall OSHIP and HHS scores reported by Barnes, and the tendency of the OSHIP scores to be somewhat lower relative to the HHS scores, suggests that the OSHIP is a very close, although *conservative*, estimate of the HHS.

#### Patient Satisfaction Data Collection Method

Patient satisfaction data was also collected using the annual, patient-completed, mail-in questionnaire. For the purpose of the BHR study, an additional question about patient satisfaction was appended to the end of the OSHIP assessment questionnaire. Patient satisfaction data was collected by presenting the patient with the following question:

Satisfaction

- ☐ I am extremely pleased with the operation.
- ☐ I am pleased with the operation.
- ☐ I am no different than before the operation.
- ☐ I am worse than before the operation.
- ☐ I am much worse and would not recommend the operation.

#### **Question #2:**

**Please discuss the evaluation methods used to collect effectiveness data (i.e., data on survivorship, OSHIP score, radiographic, and patient satisfaction) and whether or not these methods are reliable to assess the effectiveness of the device.**

**APPLICABILITY OF THE FOREIGN DATA FROM A SINGLE INVESTIGATOR AND UNITED KINGDOM PRACTICE OF MEDICINE TO THE TARGET UNITED STATES POPULATION AND PRACTICE OF MEDICINE**

**Comparison of the United States and United Kingdom Patient Populations**

The clinical data were derived from a foreign clinical study conducted by Dr. McMinn at the Birmingham Nuffield Hospital in the United Kingdom. There are no racial or ethnic origin data for the patients in the clinical study. However, the sponsor reasons that the racial and ethnic distributions in the U.S. and U.K. populations are similar and that although the percentage of the population of African-descent is higher in the U.S. than the U.K., the target U.S. population should have a significantly higher percentage of Caucasian patients. The sponsor stated that neither control group data demonstrated that their surgical patient populations reflected the ethnic distribution of the general U.S. population.

<b>Comparison of the Ethnic / Racial Distributions</b>		
	<b>U.S.</b>	<b>U.K.</b>
White	75.1%	92.1%
Black	12.3%	2.0%
Asian	3.6%	4.0%
Native American	0.9%	-
Pacific Islander	0.1%	-
Chinese	-	0.4%
Other race	5.5%	0.4%
Mixed race	2.4%	1.2%

The following table includes a description of the gender, age, and diagnostic indications for the Overall McMinn Cohort and two multi-center studies used to support recently-approved hip arthroplasty devices:

<b>Patient Demographics and Diagnostic Indication Comparisons</b>			
	<b>Overall BHR Metal/Metal Resurfacing Hip System</b>	<b>Wright Medical Ceramic/Ceramic Transcend Total Hip Replacement*</b>	<b>Howmedica Osteonics Ceramic/Ceramic Total Hip Replacement**</b>
Hips	2385		
Men	70.6% (1683)	62%	65%
Women	29.4% (702)	38%	35%
Age (range)	53.1 (13.4-86.5)	51.4	53
Age ≤65 years	91.9% (2191)	-	-
Dx: OA	75% (1789)	72.2%	78%
Dx: DDH	15.8% (377)	4.4%	-
Dx: AVN	4.1% (97)	19.7%	16%
Dx: Inflammatory	2.4% (57)	-	-
Dx: Other	2.7% (65)	-	2%
Dx: Post-Traumatic Arthritis	-	3.8%	4%

\* Data presented by sponsor from Summary of Safety and Effectiveness for Wright Medical Technology's Transcend Hip: P010001, February 3, 2003.

\*\* Data presented by sponsor taken from D'Antonio, J., Capello, W., Manley, et al., "New experience with alumina-on-alumina ceramic bearings for total hip arthroplasty," J. Arthroplasty 17(4): 390-97, 2002.

The sponsor's justification for the applicability of the foreign data to the US patient population is based on its large sample size, as well as the comparable demographics and diagnostic indications to the multi-center control group studies.

### **Description of Dr. McMinn's Practice of Medicine**

The sponsor specifies that all of the surgeries on the 2,385 cases in this PMA were performed by Dr. McMinn (with assistance from other surgeons) at the Birmingham Nuffield Hospital (except for 6 cases that were performed by Dr. McMinn at the Little Aston Hospital, Birmingham, U.K.). The sponsor stated that the practice of medicine, specifically the orthopedic practice of medicine, utilized by Dr. McMinn is the same as the standard of orthopedic practice in the U.S. The sponsor described Dr. McMinn's standard peri-operative regimen, as follows:

- Laminar air flow operating rooms with body exhaust suits
- Posterior surgical approach
- Standard surgical technique (described in the Surgical Technique Manual)
- Antibiotic prophylaxis intraoperatively and for 24 hours postoperatively (1.5g Cefuroxime)
- DVT prophylaxis using a single-dose (800 IU) intravenous heparin intraoperatively and compression stockings and low-dose aspirin postoperatively for 6 weeks
- Intraoperative venting of the femoral shaft to prevent fat/marrow emboli
- Early ambulation: full weight-bearing with a walker on postoperative day #1, progressing to crutches and canes
- Hospital discharge at postoperative day #6
- After 6 weeks postoperatively, begin range of motion exercises
- Recommended activities include swimming, pool exercise, non-impact or low-impact exercise at a gym; and, avoidance of high impact exercises during the first postoperative year

#### **Question #3:**

**Please discuss whether or not the foreign data from a single investigator and UK practice of medicine is applicable to the target US population and practice of medicine.**

## **DEVICE SAFETY**

### **Revisions**

There were 27 procedures that required revision in the Overall McMinn Cohort including 10 revisions due to a femoral neck fracture, 6 for femoral head collapse, 1 for dislocation, 2 for AVN (1 lead to femoral head collapse and 1 lead to a femoral neck fracture), and 8 for infections (2 lead to head collapse, 1 lead to a femoral neck fracture).

### **Adverse Events**

There were 2912 adverse events in the 2385 Overall McMinn Cohort including the following events:

<b>Adverse Event</b>	<b>Overall McMinn Cohort Totals</b>
Number of procedures	<b>2385</b>
Procedures with AE (%)	1669 (70%)
Total AEs	2912
AVN femoral head	35 (1%)
Femoral head collapse	15 (<1%)
Component migration/loosening	21 (<1%)
Femoral neck fracture	13 (<1%)
Impingement	3 (<1%)
Late Infection	15 (<1%)
Dislocation	8 (<1%)
Heterotopic Ossification	56 (2%)
Localized Event at implant site (clicking, etc.)	75 (3%)
Post-op Infection	41 (2%)
Wound exudate	589 (25%)
Pain	367 (15%)

### **Deaths**

There were 20 patient deaths (26 procedures) in the Overall McMinn Cohort. The sponsor stated that in no case was a death related to the BHR procedure.

### **Metal Ion Analysis**

The sponsor provided an unpublished report by Daniel J, Ziaee H, and McMinn D, entitled, “Metal ion studies in patients treated with the Birmingham Hip Resurfacing, a comparable FDA-approved device and historic metal-metal total hip replacements,” (File 4.1.5.15 on the Panel CD). The authors conducted 4 metal ion studies in patients who received BHR, Metasul metal-metal total hip replacements, and other (historic) metal-metal total hip replacements. The four studies included:

1. A short-term longitudinal study of urinary Co and Cr levels in patients with the BHR and the Metasul metal-metal total hip replacement.
2. A long-term cross-sectional study of urinary Co and Cr levels in patients with the BHR and the Metasul metal-metal total hip replacement.
3. A longitudinal study of whole blood Co and Cr levels in patients with the BHR.

4. A cross-sectional study of whole blood Co and Cr levels in patients with the BHR and metal-metal total hip replacements.

In addition, the sponsor provided a summary of 18 literature references pertaining to the medium and long-term safety of cobalt and chromium ion exposure in the subject device (BHR), metal-on-metal total hip replacements, and metal-on-polyethylene total hip replacements. These literature references are summarized in the Executive Clinical Data Summary.

**Question #4:**

**Based on the safety data in 2,385 patients in the Overall McMinn Cohort (i.e., data on revisions, adverse events, deaths) and the analysis of the metal ion literature, please discuss whether or not you believe that the data contained in this PMA provide reasonable assurance of safety?**

## **DEVICE EFFECTIVENESS**

### **Survivorship**

The survivorship estimates were based on the number of patients with no revision. The sponsor provided survivorship analyses for various cohorts and demographic subgroups. The 5-year survivorship of the 1,626 procedures in the X-ray/Oswestry combined cohort according to the Peto's adjustment method was 98.4%. The only marginally statistically significant difference in 5-year survival probability was between the patients with Osteoarthritis (98.8%) and Avascular Necrosis (92.1%) as their primary diagnostic indication.

### **Radiographic Data**

To assess the radiographic findings, the sponsor evaluated the X-Ray Cohort that consisted of the first 124 procedures performed in the series from 7/97 through 12/97, and was deemed to represent the learning curve by the sponsor. Radiographs were taken on 108 of the 118 procedures expected at 5 years post-operatively. The sponsor determined that 3 of the 108 (2.8%) patients for whom radiographs were available were radiographic failures (i.e., 1 due to acetabular lucency and a change in orientation  $>5^\circ$ , 1 due to acetabular lucency, and 1 due to femoral lucency) at 5 years.

### **Pain and Function - OSHIP Scores (unilateral procedures only)**

Of the 1626 procedures in the X-Ray/Oswestry Combined Cohort, 1111 were unilateral and were evaluated for effectiveness using the OSHIP questionnaire. At 5-years, 395 unilateral procedures were theoretically due and 360 procedures (91.1%) were evaluated. The mean OSHIP Scores for unilateral procedures in the X-Ray/Oswestry Combined Cohort improved from a baseline mean of 60.1 to 94.8 at 5 years. For the group of patients who had high baseline OSHIP scores ( $\geq 80$ ), the mean OSHIP scores improved from 84.5 to 99.3. The group of patients who had low baseline OSHIP scores ( $< 80$ ), the mean OSHIP scores improved from 59.4 to 95.6. At postoperative years 2, 3, 4 and 5,

the percentage of cases with good or excellent OSHIP scores was 96.9%, 95.8%, 95.2%, and 92.8%, respectively.

### **Patient Satisfaction**

The patient satisfaction is not a standard component of the OSHIP assessment but was an additional question asked for this study. At 5 years, 99.5% of all procedures in the X-Ray/Oswestry combined cohort and 99.2% of the unilateral procedures in the X-Ray/Oswestry combined cohort were pleased or very pleased with the operation.

#### **Question #5:**

##### **Based on the:**

- **5-year survivorship analysis of the 1,626 procedures in the X-ray/Oswestry combined cohort;**
- **5-year radiographic data of the 124 procedures in the X-Ray cohort;**
- **5-year pain and function (OSHIP) data of the 1,111 unilateral procedures in the X-Ray/ Oswestry combined cohort; and**
- **5-year patient satisfaction analysis of the 1,626 procedures in the X-Ray/Oswestry combined cohort;**

**Please discuss whether or not you believe that the data contained in this PMA provide reasonable assurance of effectiveness?**



## **LABELING**

Generally, prospective clinical investigations pre-define the study population with specific inclusion and exclusion criteria. This theoretically allows the study results to be generalized to that diagnostic group. In case series studies, it is more difficult to generalize the results to a defined population because the patients were not enrolled for pre-defined conditions. This is the case for the clinical data provided in this PMA submission. The clinical data were derived from the surgical experience of a single surgeon who used the Birmingham Resurfacing Hip System. The sponsor did not identify the set of diagnostic indications for the device, but instead provided a list of the diagnoses for the patients implanted with the device.

Because these patients were not investigated under an investigational protocol with pre-defined inclusion and exclusion criteria, it may be important to determine the diagnoses of all of the patients who were evaluated by Dr. McMinn during this same time period but did not receive the Birmingham Resurfacing Hip device and either had no surgery or had surgery with a conventional total hip replacement or other device. However, a complete review of these patients was not presented in the PMA. With this information, it might have been possible to retrospectively determine what criteria, if any, were used to select candidates for the Birmingham Resurfacing Hip device. However, to retrospectively develop the indications for use and physician labeling, the sponsor provided a list of the factors that contributed to Dr. McMinn's decision to perform a total hip replacement (THR) in certain patients rather than a hip resurfacing procedure (BHR). These factors included:

- Advanced age: Patients of advanced age, especially those with low activity levels, were typically candidates for THR rather than BHR. Only 8.1% of the 2,385 cases included in the Overall McMinn cohort were >65 years of age. In these cases, BHR was selected despite advanced age if the patients had high activity levels, and had good bone stock of the femoral head.
- Low activity level: Patients with a low activity level were considered at lowered risk for future revision, and therefore good candidates for THR. Low activity level was characterized by no participation in sports activities, no heavy work required by job, a sedentary/retired lifestyle, or comorbidities that precluded a high activity level, such as severe arthritis in other joints or severe heart disease.
- Poor bone stock: Patient with poor bone stock were selected for THR rather than BHR because they were considered at risk for femoral neck fracture or femoral head collapse with a hip resurfacing procedure. Poor bone stock was characterized as severe osteopenia of the femoral head or femoral neck (determined by risk factors, medical history and/or diagnostic imaging), extensive AVN (>50% of femoral head, regardless of FICAT Grade), or the presence of multiple cysts.

The sponsor stated that Dr. McMinn's collection of a patient's pre-operative history, physical, and diagnostic work-up was commonly sufficient to screen candidates for BHR versus THR, and that only in rare instances would the planned surgical procedure be revised intraoperative. The sponsor stated that although Dr. McMinn rarely changed his

preoperative plan based on intraoperative findings, all patients were consented for a hip arthroplasty, and informed about the probable type of prosthesis they would receive. As with any surgical procedure, patients were also informed that based on the intraoperative findings, there could be changes to the planned procedure. The patients were thus consented for both a BHR and THR procedure.

Based upon the patient population studied, the factors outline above, and an analysis of the BHR revisions in the Overall McMinn Cohort (i.e., femoral neck fracture, femoral head collapse, dislocation, AVN, and infection), the sponsor proposed the following indications for use and contraindications for the device:

### **Indications for Use**

The Birmingham Hip Resurfacing (BHR) System is a single use device intended for hybrid fixation: cemented femoral head component and cementless acetabular component. The BHR system is intended for use in patients requiring primary hip resurfacing arthroplasty due to:

- Non-inflammatory arthritis (degenerative joint disease) such as osteoarthritis, traumatic arthritis, avascular necrosis, or dysplasia/DDH, or
- Inflammatory arthritis such as rheumatoid arthritis.

BHR System hip resurfacing arthroplasty is intended for joint replacement in patients who are at risk of requiring future, ipsilateral hip joint revision. While it is impossible to predict if a patient will require more than one joint replacement, several factors are known to increase risk of revision surgery including age less than 55 years at index surgery and/or high physical activity level postoperative.

### **Contraindications**

Contraindications for use of the Birmingham Hip Resurfacing (BHR) System include:

- Patients with infection or sepsis,
- Patients who are skeletally immature,
- Patients with any vascular insufficiency, muscular atrophy, or neuromuscular disease severe enough to compromise implant stability or postoperative recovery,
- Patients with bone stock inadequate to support the device including:
  - Patients with severe osteopenia should not receive a BHR procedure. Patients with a family history of severe osteoporosis or severe osteopenia.
  - Patients with osteonecrosis or avascular necrosis (AVN) with >50% involvement of the femoral head (regardless of FICAT Grade) should not receive a BHR.
  - Patients with multiple cysts of the femoral head (>1cm) should not receive a BHR.
- Females of child-bearing age due to unknown effect on the fetus of metal ion release.
- Patients with known moderate to severe renal insufficiency.

In addition to the factors described above, the sponsor also considered the review of 50 BHR femoral neck fractures reported by Shimmin and Back<sup>7</sup> in their development of the

---

<sup>7</sup> Shimmin AJ, Back D. Femoral neck fractures following Birmingham hip resurfacing: A national review of 50 cases. J Bone Joint Surg [Br] 87-B:463-4, 2005.

labeling. In this publication, the authors reported a review of 3497 BHR cases performed in Australia by 89 surgeons. There were 50 femoral neck fractures in the series (1.46%) which the authors attributed to osteoporosis, notching of the superior femoral neck, varus placement of the device by more than 5°, and technical difficulties including poor exposure due to obesity, change in intra-operative alignment, and poor impaction of the femoral component. Based on these findings, the sponsor added the following warnings and precautions to the labeling:

**Warnings**

- Avoid notching the femoral neck, as this may lead to femoral neck fracture.
- Avoid placing the femoral component in varus. Varus placement of the femoral component has been associated with femoral neck fracture.

**Precautions**

- Improper selection, placement, positioning, and fixation of the implant component may result in early implant failure.

**Question #6:**

**Do the patient selection methods and data presented on the BHR device support the proposed labeling indication?**

**Please comment on any other aspects of the product labeling, such as:**

- **Contraindications**
- **Warnings**
- **Precautions**
- **Potential Adverse Effects on Health**

## **POST-APPROVAL STUDY**

The FDA advised the sponsor that the PMA may be subject to conditions of approval including a post-approval study to evaluate the long-term safety and effectiveness of the device. In response to FDA's advisory, the sponsor included a post-approval study protocol.

The proposed post-approval study is a prospective, non-randomized, longitudinal, unblinded, multicenter trial to evaluate the long-term safety and effectiveness of the device. The sponsor proposes to enroll 150 patients at up to 15 investigational sites who meet the inclusion and exclusion criteria and sign the informed consent. Patients will be clinically and radiographically evaluated preoperatively, intraoperatively, and postoperatively at 6 months, 1 year, 2 years, 3 years, 4 years, and 5 years. Continued long-term follow-up assessments will be performed using a self-administered, mail-in patient questionnaire from 6 years to 10 years. Explanted device components will be analyzed according to an explant protocol. Clinical and radiographic success and failure criteria were defined.

### **Question #7**

**A reasonable assurance of safety and effectiveness as defined in questions #4 and #5 above must be demonstrated for device approval. If you believe the data in the PMA demonstrate a reasonable assurance of safety and effectiveness but think there are remaining specific questions regarding this device that should be addressed in a post-approval study, please identify those questions.**